



---

# DELIVERABLE

**Project Acronym:** DM2E

**Grant Agreement number:** ICT-PSP-297274

**Project Title:** Digitised Manuscripts to Europeana

---

## D5.10 Results Transfer Plan

---

**Revision:** Final 1.0

---

**Authors:**

Doron Goldfarb (ONB)  
Kai Eckert (UMA)  
Antoine Isaac (Europeana)  
Christian Morbidoni (Net7)  
Lieke Ploeger (OKFN)  
Vivien Petras (UBER)  
Violeta Trkulja (UBER)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PP	Restricted to other programme participants (including the Commission Services)	x

## Revision history and statement of originality

Revision	Date	Author	Organisation	Description
0.1	18.06.14	VT	UBER	Initial Draft
0.2	03.07.14	DG, KE, CM, LP, VP, VT	ONB, UMA, Net7, OKFN, UBER	Added content to all sections
0.3	04.07.14	VP, VT	UBER	Corrections by UBER
0.4	14.07.14	AI, DG, KE, CM, LP, VP, VT	Europeana, ONB, UMA, Net7, OKFN, UBER	Expanded draft
0.5	24.07.14	VT	UBER	Revision of all sections and addition of content to section 9
0.6	26.07.14	VP	UBER	Final Revision
0.7	28.07.14	VT	UBER	Input on section 10
0.8	29.07.14	DG, VT	ONB, UBER	Revision of section 4.1, 7.3 and 9
Final 1.0	30.07.14	VP	UBER	Approval Final 1.0

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

---

## Contents

<b>1</b>	<b>Sustainability of Project Results - Original Plan and Deviation from the DoW...</b>	<b>5</b>
<b>2</b>	<b>DM2E Achievements .....</b>	<b>7</b>
<b>3</b>	<b>Potential Cooperation Partners .....</b>	<b>9</b>
<b>4</b>	<b>DM2E Data.....</b>	<b>11</b>
4.1	Object Metadata.....	11
4.1.1	Providing DM2E Object Metadata to Europeana.....	11
4.1.2	Archiving the Object Metadata.....	11
4.1.3	Maintaining the DM2E Server .....	12
4.1.4	Future Data Updates / Finding alternative aggregators.....	12
4.2	Annotation Data.....	12
4.2.1	Annotation Storage.....	13
4.2.2	Open Annotation .....	13
4.2.3	Maintenance of the Public Web Applications and Services .....	13
4.2.4	Independent Controlled Deployments .....	14
4.3	Europeana Channels .....	14
<b>5</b>	<b>Infrastructure .....</b>	<b>15</b>
5.1	Developed Infrastructure (OmNom) .....	15
5.1.1	Software Storage .....	15
5.1.2	Server Maintenance .....	16
5.1.3	Ingested Data and the Linked Open Data API .....	16
5.2	Scholarly Research Platform (Pundit) .....	16
5.2.1	Net7 Maintains Software Development .....	16
5.2.2	Software Storage .....	16
5.2.3	Integration with Europeana.....	17
<b>6</b>	<b>DM2E Model.....</b>	<b>18</b>
6.1	The DM2E Ontology .....	18
6.2	RDF Application Profiles.....	18
<b>7</b>	<b>Documentation .....</b>	<b>19</b>
7.1	Europeana Pro .....	19
7.2	Deliverables .....	19
7.3	DM2E Wiki.....	19
7.4	Mapping Guidelines .....	21
7.5	DM2E Website .....	21
7.6	Project Management Platform .....	22

---

7.7	Github .....	22
7.7.1	Ontologies.....	22
7.7.2	Scholarly research platform "Pundit" .....	22
7.7.3	OmNom .....	22
7.7.4	Validation Tools.....	23
7.8	Direct Ingestion Tools .....	23
7.9	DM2E Analyses .....	23
7.10	Pubby23 .....	
7.11	Vimeo.....	24
<b>8</b>	<b>Community.....</b>	<b>25</b>
8.1	Open Glam .....	25
8.2	Open Humanities Awards.....	25
<b>9</b>	<b>Appendix A: DM2E Proposal for Extending the EDM Implementation of Europeana .....</b>	<b>27</b>
9.1	Representing DM2E annotatable content in EDM .....	28
9.1.1	Introduction .....	28
9.1.2	Providing DM2E Object Metadata to Europeana.....	29
9.1.3	Alternative Representation of Low-level, Annotatable Content (= Pages) .....	29
9.1.3.1	Disadvantages of Current EDM Implementation .....	30
9.1.3.2	Proposal for two EDM Extensions.....	30
9.1.3.3	Discussion .....	31
9.2	Providing a Link to the Annotation Software .....	32
9.2.1	Introduction .....	32
9.2.2	Providing means to link from the Europeana portal to external representations .	32
9.2.2.1	Disadvantages of current options to provide additional external links.....	33
9.2.2.2	Proposed Solution .....	33
9.2.2.3	Discussion .....	34
9.3	Necessary Changes for the DM2E Annotation Service .....	35
9.4	Current state of discussion with Europeana .....	36
9.4.1	Outcome of Discussion between Europeana Product Development and R&D regarding proposed changes section 9.1.3 .....	36
9.4.2	Outcome of Discussion between Europeana Product Development and R&D regarding proposed changes section 9.2.2 .....	36
<b>10</b>	<b>Appendix B: Follow-Up Aggregators.....</b>	<b>37</b>

## 1 Sustainability of Project Results - Original Plan and Deviation from the DoW

The original plan for the sustainability of project results was (from the DoW, Task 5.6):

*“In order to guarantee reuse of the project results and thus their sustainability, HUB, the Europeana Foundation and TEL will work together on a sustainability plan. The overall objective is not so much to create a new organization at the end of the project but to make sure that data, technical building blocks and platform infrastructure created in DM2E will be transferred to Europeana/The European Library to be hosted and reused. Input from the leaders of WPs 2, 3 and 4 is required for this plan. The outcome of this task is twofold: a results transfer plan will need to be created roughly 6 months before the end of the project and the actual transfer of results needs to take place at the end of the project.”*

The DM2E project started in September 2013 to think about the sustainability of the project results, more than a year before the project was scheduled to end. In that early stage it became clear that Europeana will not be able to follow the plan as laid out in task 5.6.

As a matter of record, DM2E would like to state that there are no technical or other barriers from the DM2E side to transfer data and infrastructure to Europeana, most of which resides on the DM2E server, it is rather a problem of Europeana not being equipped to process the project outcomes.

When DM2E started, Europeana had already started the Linked Open Data Pilot and was planning the ingestion of RDF data. The idea of ingesting data based on specialisations to the EDM model (like the DM2E model) and handling their representation in the Europeana portal was also discussed as a plan and appeared as viable options for DM2E's data and infrastructure to be implemented. Additionally, annotation - as provided by Pundit has always been a Europeana business goal. This means that none of the developments in the DM2E project have been done in vain, it is rather a challenge of Europeana's development processes catching up with the approaches tested in DM2E's plans.

Europeana states:

*“Europeana cannot realise the transfer of DM2E's infrastructure into its core infrastructure and data because of the difficulty to handle different storage systems at once, and different flavours of data. DM2E uses a flexible RDF store that fits perfectly its application goals. Using such a store in production for Europeana is too premature. Europeana has moved towards an entirely new infrastructure not so long ago and lacked the time to experiment further with RDF storage. Projects like DM2E help prototyping such move, but their scale and the diversity of involved partners is not the same as Europeana's. The data workflows created in DM2E, while comparable with Europeana's, are still different and would require adaptation efforts to fit Europeana's current data flows, which have been themselves changed in the past months (for example on metadata enrichment and indexing). Handling different levels of data (as would be the case for data expressed according to the specialized DM2E model on the one hand and the 'basic' EDM model on the other hand) is also impossible now. For the implementation of the Europeana Channels though (see section 4.3) solutions will have to be found to have extensions to EDM co-exist, which may represent a step towards being able to handle better DM2E data in Europeana's core services.” (Antoine Isaac, on behalf of Europeana, 07/2013).*



---

This "D5.10 Results Transfer Plan" first introduces the achievements of the created Linked Data infrastructure within the DM2E project and then in section 3 describes the conducted contacts of our project to potential cooperation partners from the Europeana cluster and representatives from other institutions in order to sustain our project results. In the course of the DM2E project, object metadata was delivered by the content providers and annotation data was generated by users of the provided content. Archiving, access and integration with Europeana are described in section 4. Section 5 explains where and how the developed infrastructure will be sustained in order to be available in the future. The DM2E model, a specialisation of the Europeana Data Model (EDM) and the initiated task group on RDF Application Profiles at the Dublin Core Metadata Initiative are explained in section 6. Section 7 describes where deliverables, the DM2E wiki, software, data model and learning videos will be documented. This document concludes with the support of the OKFN on the established network around the DM2E. The Appendix A proposes extensions to Europeana's EDM implementation based on requirements identified throughout the course of the DM2E project. The underlying rationale is to transfer the DM2E specific metadata representation of annotatable text/image content into a form that can be hosted by Europeana without relying on specialised DM2E data infrastructure and to make the data available for annotation.

---

## 2 DM2E Achievements

DM2E laid the ground for important, further developments. Looking at the components, there is first and foremost the DM2E model, a very rich data model that reflects the requirements and views of many data providers and scholars regarding the fine-grained representation of digitised manuscripts. The DM2E model was developed as a pure specialisation of EDM and therefore is an ideal example of an application profile that supports a specific application while retaining the general interpretability and compatibility with Europeana.

Next, DM2E created an infrastructure to provide Linked Data representations of digitised manuscripts (or any other resource) in a stable way, using versioning and provenance information. Again, a main achievement is to be seen in the underlying data model and the general architecture that has been created. All components that have been developed can be used and reused independently from each other:

- DM2E compatible data can be created either by using MINT, if a graphical editor is preferred, or by adapting one of the existing XSLT mappings that are documented in the DM2E Wiki (see section 7.3). For a single data provider, the direct ingestion tool is the easiest way to transform and ingest own data into a triple store, with versioning and provenance support.
- OmNom and the DM2E workflow engine is more difficult to set up, but this software is targeted at aggregators or service providers, not single data providers. Nevertheless, with such a system, we demonstrated how workflows can be executed on the Web, distributed, if desired, but most importantly, in a transparent way with a full provenance chain linked to the resulting data.
- DM2E data finally can be provided using DM2E Pubby, a very advanced version of the established Pubby software for Linked Data publishing that is nonetheless simple to deploy and to configure. Backed by a Fuseki triple store, the setup of a new Linked Data site can be done in short time even for non-experts.
- Finally, the provided data can be used to directly work with the manuscripts using tools like Pundit. Considering a scenario where Europeana provides rich data like DM2E data to support various applications from third-parties, Pundit would be such a third-party solution. DM2E demonstrated how such an open data environment can actually work.
- The deployed DM2E infrastructure, in the sense of a “living” platform, cannot be fully sustainable without additional funding, despite any efforts UBER is willing to make to keep the servers running. However, everything is in place to reach full sustainability: Linked Data can and should be distributed. Ideally, the data providers should provide access to their data directly. Only a search facility is needed to get access to the distributed data, which can be provided by any third-party (DM2E demonstrated this as well with a commercial search product and an additional open-source solution) or, of course, by Europeana.

If data providers cannot provide the data because of licensing issues – which is not easily acceptable at least in the long run – or if redundancy or additional stability of the data is to be provided, then reliable institutions are needed to function as a hub and service provider. This can be Europeana directly or a more distributed network of aggregators providing additional services beyond mere data ingestion.



---

This might be a vision, but the vision is in line with actual development plans at Europeana, notably Europeana Cloud. DM2E provides an important prototype for these developments.

The results transfer plan in this document details ideas, approaches and solutions the DM2E project partners devised in order to provide alternatives to the original plan that would still enable the sustainability of the DM2E project results by other means than transferring both the data, technical building blocks and platform infrastructure to Europeana / The European Library.

### 3 Potential Cooperation Partners

While the results transfer plan has been a frequent topic at DM2E executive board meetings, sustainability issues and options were discussed with all partners at DM2E's All-WP-Meeting-4 in Athens (Nov. 2013) and All-WP-Meeting-5 in Bergen (June 2014) as well.

At an early stage, the project also started to conduct conversations with

- Europeana Cloud
- Europeana Creative
- Europeana Inside

of which Europeana Cloud was discussed as a candidate project for taking over selected infrastructure components, particularly the annotation platform.

DM2E also explored other possibilities to sustain its data and infrastructure. The project coordinator established contact with representatives from:

- DARIAH<sup>1</sup>
- Textgrid<sup>2</sup>
- Computer and Media Service of Humboldt-Universität zu Berlin<sup>3</sup>

to talk about possible hosting and transfer of our project results to long-term available options. While DARIAH expects their contributors to host their own infrastructure, Textgrid offers a fee-based service to host the data. Humboldt's Computer and Media Service's can also host the data, but not the software infrastructure. In both cases, where data can be hosted, the options that were discussed were based on static XML-formatted dumps of the data, a minimal solution for long-term preservation. None of the projects and partners contacted had means to host a linked data based infrastructure like DM2E's.

While this makes transferring our technology more difficult, we also regard this as a signal for the innovativeness of our linked data approach. We believe that the consistent use of linked data technology is a step forward to increased interoperability between different data providers and information systems. We plan to continue with this line of development in a follow up project which will be submitted as a proposal in the EU's Horizon 2020 programme. The intended project is dedicated to the challenges of the creation of virtual research infrastructure for scholars and will be based on the results of DM2E as well on standards and solutions from the Semantic Web and the Linked Open Data (LOD) Web. One of the main project objectives will be also to discover innovative and stable long-term preservation strategy for the sustainability of data.

During the project, we have been in constant contact with Europeana: a Europeana representative attended all project meetings as well as a number of executive board meetings. Europeana is working on publishing Project Shift Guidelines, which serve to guide projects after the funding period ends. DM2E has commented on a draft of the document. Some of the suggested alternatives are presented in this document.

DM2E could identify the type of the project (scenario 1 - project ends). It conducted a survey to collect all relevant information for data provider shifts to other aggregators of

---

<sup>1</sup> <https://www.dariah.eu/>

<sup>2</sup> <http://www.textgrid.de/en/>

<sup>3</sup> <http://www.cms.hu-berlin.de/>



---

the project partners and shared the survey results with Europeana. Europeana is assisting DM2E in finding aggregators for the DM2E providers (see section 10, Appendix B).

---

## 4 DM2E Data

### 4.1 Object Metadata

The DM2E project will publish its data to Europeana during DM2E's lifespan, using Europeana's existing ingestion process. As this is already covered in the subcontract between DM2E and the Europeana Foundation, updates will be delivered during the duration of the project by DM2E as aggregator for all data providers in the framework of the DM2E project.

#### 4.1.1 Providing DM2E Object Metadata to Europeana

The DM2E project aggregates data from providers using an extension of the EDM, the DM2E data model. Europeana is not able to aggregate this extension of the EDM yet and will therefore ingest the 'basic' EDM. DM2E will convert its RDF data to fit the constraints set in EDM/XML and will transfer it to Europeana via OAI-PMH. One significant feature of DM2E's contribution, however, is making the digitised content behind some of the provided metadata records accessible to annotation software such as the tool Pundit which is being developed as part of the DM2E project. Through converting DM2E metadata from the DM2E data model to EDM, this feature currently gets lost.

DM2E therefore developed a proposal seeking to identify possibilities to represent annotatable DM2E content in EDM in a form that

1. is compatible with EDM
2. fulfils Europeana's requirements regarding the granularity of provided objects
3. is compatible with DM2E's annotation software Pundit.

Section 9 (Appendix A) presents the proposal for extending the EDM implementation of Europeana. It describes current metadata representation of DM2E annotatable content within the project's data infrastructure and discusses issues with providing the same functionality in Europeana's EDM implementation.

#### 4.1.2 Archiving the Object Metadata

As the Europeana representation of the DM2E data will differ from the project's original metadata format, it will be necessary to preserve DM2E's output for future use. The minimal solution will therefore be to archive all data for later retrieval.

##### **Complete data archive**

DM2E will provide a complete archive of the data on the server hosted by the Computer and Media Service Repository of Humboldt-Universität. This way it will be ensured that no data gets lost and that the data will be accessible in future.

##### **Data dump of the Object Metadata to all Content Providers**

In addition to offering a complete archive of all the metadata, the project will also provide separate dumps for each data provider with their data in both DM2E and EDM format.

---

### 4.1.3 Maintaining the DM2E Server

In order to maintain the annotation functionality based on the DM2E data representation, the DM2E data infrastructure must be kept running. An extended solution would therefore be to provide access to the data via the DM2E server as long as possible. Since the DM2E infrastructure is already located at Humboldt, this should be straightforward as long as no hardware/software failures occur. It should be pointed out that the annotation platform can also run on data not provided in the DM2E format, so if the DM2E data server would be down, the Pundit software is still able to work on other data.

### 4.1.4 Future Data Updates / Finding alternative aggregators

The current Europeana aggregation approach requires aggregators. The DM2E project cannot act as an aggregator once the project ends as such a sustainable incorporation as long-term organisation was not planned for DM2E. In order to support future updates of the data provided to DM2E and continued cooperation of DM2E's providers with Europeana, other aggregators shall take up DM2E's role with respect to the DM2E data providers. Europeana is committed to give the DM2E project advice in this respect.

Europeana will check all DM2E providers, to see where they could be associated with a current Europeana aggregator (some DM2E providers are already sending data to other aggregators than DM2E). The DM2E providers themselves can also decide to which aggregator they would like to go, once the DM2E project ends.

One possible option for an aggregator is *The European Library*. The DM2E providers could join TEL if they are able to bear the costs for this service. Depending on the outcome of DM2E's proposal regarding annotatable content it would be of high priority to identify aggregators that are able to ingest data delivered as EDM. Europeana is assisting DM2E in finding aggregators for the DM2E providers (see section 10, Appendix B).

Currently the DM2E data infrastructure implements the transformation from DM2E to EDM in a "centralised" manner. First, providers ingest their data to DM2E. In a separate step, the DM2E data is then converted to EDM. In order to allow providers to independently use their existing DM2E mappings for further EDM conversion, DM2E will provide a standalone tool chain that accepts legacy metadata and the existing legacy->DM2E XSLT mappings and outputs the metadata converted to EDM.

## 4.2 Annotation Data

The annotations produced by the annotation platform *Pundit*, which reflect the work from researchers with the manuscripts from the DM2E providers, are an important outcome of the DM2E project. Right now these kind of annotations, represented in the form of RDF using a subset of the Open Annotation model, cannot be ingested into the Europeana platform and this won't change before the end of the DM2E project. On the other hand, scholars' annotations are a sort of "personal data" that can be private for the use of a single scholar, or shared. This means that transferring those annotations to Europeana would imply the existence of some collaborative annotation functionality in Europeana itself (in order to be able to effectively use such data).

Europeana recently scheduled the development of annotation functions (in Q4 2014). Even if the envisioned annotation facility will be simpler (i.e. a subset of annotations facilities found in Pundit), the use of RDF and Open Annotation will make it possible to

---

properly transform and re-use the annotations produced in Pundit, in the case they are of interest to Europeana (See section 4.2.2 for more details).

The sustainability actions described in the following will ensure that scholars will be able to access the annotation tools in the future without losing their annotations.

#### 4.2.1 Annotation Storage

Humboldt-Universität zu Berlin will archive the existing annotation data (a snapshot of all the annotations created during DM2E) as a package in the repository at the Computer and Media Service. Additionally, the source code of the scholarly research platform tools (updated to the last version produced within the project), along with all the libraries, tools and documentation needed to set them up, will be stored on the same server. This will ensure that, even in the case Net7 will stop maintaining the tools, new deployments of the tools will be possible.

As annotations in Pundit can be private and public, in order to ensure privacy of the private annotations, the snapshot will include only public annotations created in the frame of the DM2E project.

In the case the Net7 public server will be dismissed in the future, users will be notified and action will be required from them to allow export and storage of private annotations.

#### 4.2.2 Open Annotation

All Pundit annotations are exportable via the Pundit server API. Public annotations are served via an open API, while private annotation needs authenticated APIs (using standard OpenID workflow).

Annotations are represented in Open Annotation and use named graphs as described in the [the AO extension specification \(<http://www.openannotation.org/spec/extension/#NamedGraph>\)](http://www.openannotation.org/spec/extension/#NamedGraph).

The advantage of such a compliance makes the Pundit annotations independent from EDM and from a specific annotation tool (Pundit):

- annotations are independent from DM2E provider data
- Pundit will have an export functionality to map to Open Annotation
- Third party applications that supports Open Annotation will be able to ingest annotation created with Pundit and make use of them.

Europeana's annotation functions will also be based on Open Annotation, and will adopt specific EDM/OA data patterns that will also be used in other projects (EuropeanaCreative, Europeana 1914-1918, Europeana v2.0). This means that, in order to reuse Pundit annotations, some transformation would be needed to adhere to the data patterns that Europeana will choose to adopt.

#### 4.2.3 Maintenance of the Public Web Applications and Services

The core components of the DM2E scholarly research platform, namely, Pundit, Ask, Feed and Korbo, are one of the main business assets at Net7. There is no intention to stop

---

maintenance and development of the core components and the relative open web services (<http://ask.thepund.it>, <http://feed.thepund.it> and <http://as.thepund.it>) in the next years.

However, Net7 will ensure to copy the latest version of all the tools (including source code, libraries and documentation) as well as the annotation data to the Computer and Media Service server at Humboldt Universität, in the case a different decision will be taken in the future to stop maintaining the services. In this way new deployments of each tool will be possible, provided that they will adhere to the non-commercial license.

The Pundit client and Ask are open-source with an AGPL3 license. The Pundit server has a dual license that ensures the release of the code as AGPL3 after a maximum of 6 months. Net7 will ensure that the latest version of the Pundit server produced in the frame of the DM2E project will be available as AGPL3 before the end of the DM2E project. The other two core components (Feed and Korbo) are currently under code revision and will be published with the same licensing policy used for the Pundit client.

#### 4.2.4 Independent Controlled Deployments

Projects like Wittgenstein and Burckhardt and corresponding user communities can decide to run their own server with a Pundit installation and store their own annotations on their own server to have the full control on the annotation data in the future.

### 4.3 Europeana Channels

In the context of the EuropeanaSounds project, Europeana will develop a platform for creating channels<sup>4</sup>. These channels will consist of segmented sub-portals curated and designed with the Europeana network. These will be tailored to fit the needs of specific content themes. Objects from Europeana will be thematically grouped, contextualised and marketed to specific user audiences (for example food and drink, sound, fashion etc.). The goal is especially to provide users with selected collections, improved discovery (e.g. facets w/ domain-specific vocabularies), direct access to media via specific viewers, and the ability for users to annotate metadata records and media.

To implement the channels, it is envisioned that the platform will be able to handle (domain-) specific extensions to the standard EDM classes and properties. If the DM2E partners were able to create a specific channel for the DM2E content, this could enable to have the richer DM2E data be ingested and employed to provide finer-grained search, display and annotation services, better aligned with the tools and results of DM2E project.

DM2E is very interested in developing such a channel and is in conversation with Europeana for a Digital Humanities Channel. Additional resources are needed for this endeavor.

---

<sup>4</sup> <http://de.slideshare.net/DavidHaskiya/e-sounds-kickoffdavidhaskiya2>

---

## 5 Infrastructure

Europeana will not have the resources to maintain services like OmNom and Pundit; they will not become internal services features in Europeana's own product development.

### 5.1 Developed Infrastructure (OmNom)

As outlined above, ideally, Europeana would provide services that make DM2E's infrastructure obsolete. These services can be created using the software developed in DM2E, but this is not essential. The software can also be seen as a prototype implementation and the question to what extent it can be reused directly is to be decided by Europeana developers.

In the remainder of this section, it has to be distinguished between

1. the provision of the ingestion system to fulfil the role as an aggregator by ingesting new data, possibly from new data providers,
2. the provision of the developed infrastructure as stand-alone software to be reused by data providers, and
3. the provision of the ingested data to be accessible by scholars using "linked data aware" tools like the one implemented in the DM2E scholarly research platform.

The first can only be done with human resources, as the infrastructure not only has to be maintained, but also actively promoted and supported to be usable by data providers. Creating rich DM2E metadata and in particular ingesting it to Europeana is not and never will be a self-explanatory task like for instance a webmail system.

The second will be done. In particular we will document how a DM2E compatible access to digitised manuscripts can be realised with minimal efforts.

The third is absolutely crucial and will be in the focus of this sustainability plan. At least for the time until Europeana is able to host rich application-specific metadata, the data and the access to it has to be preserved.

#### 5.1.1 Software Storage

The software will be provided as binary and source packages to be downloaded from the DM2E server hosted at UBER. There are the following packages that can be deployed independently:

1. MINT to create mappings to DM2E visually
2. XSLT mappings with documentation as templates for custom XSLT mappings (XSLT mappings are on Wiki documentation)
3. The DM2E workflow engine for fully-distributable, Web-based data transformation and ingestion workflows, including single-sign-on support.
4. OmNom as graphical user interface to the workflow engine.
5. Direct Ingest, a command line tool that is better suited for single data providers who do not want to provide a full workflow system, but rather just ingest their own data.
6. Silk and the Silk workbench to contextualise the data.

- 
7. DM2E Pubby to create the Linked Data API to be accessed by consuming applications like Feed (feed.thepund.it).

As a quick-start solution and for tests, virtual machine images will be provided to start the full infrastructure locally with minimal effort.

The Git repositories will remain on Github for convenient access by developers, but also cloned and provided via the DM2E server as backup.

### 5.1.2 Server Maintenance

The whole infrastructure is deployed at the dedicated DM2E server hosted at UBER. Currently, Ubuntu Server 12.04 LTS is installed, with guaranteed updates until 2017.<sup>5</sup> At the end of the project, DM2E will update the operating system to the current LTS version 14.04 with support until April 2019. Humboldt-Universität commits to maintain the server at least for this time period, including necessary hardware replacements. After 2019, the server will be kept online as long as possible, depending on the hardware and the ability to apply security updates to the software.

### 5.1.3 Ingested Data and the Linked Open Data API

As long as the server is maintained, the ingested data will be accessible via the Linked Data API. This provides a quite long timeframe during which hopefully more sustainable and flexible solutions within the Europeana ecosystem can be developed. As stated above, we envision two scenarios that ensure real sustainability: either a distributed provision of Linked Data by the data providers or a flexible data provision of application-specific rich EDM data like DM2E data by Europeana.

As long as the server will be maintained/hosted by Humboldt-Universität, the LOD API will be accessible.

## 5.2 Scholarly Research Platform (Pundit)

### 5.2.1 Net7 Maintains Software Development

Net7 is actively using the tools Pundit, Korbo, Ask and Feed in other R&D and commercial projects. Net7's plan for the future is to continue developing and improving the tools independently from the DM2E project. All tools will be available in GitHub and released under a dual licence (A-GPL 3 + commercial license as described in <http://thepund.it/license>). Net7 will continue maintaining a public instance of the Pundit annotation server where users can store and retrieve their annotations.

### 5.2.2 Software Storage

There is no intention to stop maintenance and development in the next years, however Net7 will ensure to copy the latest version of all the tools (including source code, libraries

---

<sup>5</sup> <https://wiki.ubuntu.com/LTS>

---

and documentation) as well as the annotation data to the Computer and Media Service server at Humboldt-Universität, in the case a different decision will be taken in the future to stop maintaining the services. In this way new deployments of each tool will be possible, provided that they will adhere to the non-commercial license.

### 5.2.3 Integration with Europeana

In the coming months, a complex annotation suite like Pundit cannot be re-used directly for annotation on the core Europeana services. However, it may be connected with dedicated, annotation-intensive Europeana Channels (see section 4.3). Pundit could also be developed to be aligned with the services EuropeanaCloud will develop for digitised content. Contrary to Europeana, EuropeanaCloud will work on hosting digitised objects. This offers great opportunities to demonstrate Pundit's potential value on the longer term.

Requirements for content to be annotated and for Linked Data representations to support redirection to the annotations environment have been provided as result of WP1 (see section 9). The same rules could be applied to integrate Pundit with the Europeana Cloud, portal or channels.

Minimum requirement to annotate HTML with Pundit is that such HTML should not include sophisticated Javascript applications (e.g. image viewer) or other kinds of embedded applications (e.g. flash based, Java Applets, etc.), and that the images that one wants to annotate are provided as simple IMG HTML tags (not specific image viewers).

---

## 6 DM2E Model

### 6.1 The DM2E Ontology

The DM2E model is the current specialisation of the Europeana Data Model (EDM) for the object metadata within the DM2E project. It is an ontology that has been developed regarding the requirements of the projects' data providers and is fully based on the Linked Data paradigm. It thus allows the easy distribution of the ontology beyond the projects' end via different ontology repositories. The last version of the model which will be released until the end of the project will be distributed twofold in order to achieve higher visibility and sustainability of the model. Apart from DM2E's own ontology publishing service<sup>6</sup>, an installation of the Web-based RDF Schema vocabulary publishing system Neologism, the model will be also published on external ontology libraries like: (<http://lov.okfn.org/>, <http://vocab.deri.ie/>, <http://metadataregistry.org/>, <http://datahub.io/>, <https://onki.fi/>). As long as the DM2E server will be maintained by UBER's Computer- and Media Service, the ontology will be further available by supporting content negotiation. Furthermore the ontology will be described by extended metadata about versioning, modifications and provenance.

### 6.2 RDF Application Profiles

An important contribution of DM2E is the fundamental work towards EDM application profiles<sup>7</sup>, where DM2E provides an application-specific, richer view on the metadata. Questions to be answered are how to express and represent the relationship between EDM and its application profiles like DM2E in the data model, how to express and represent constraints on the data within the data model and how to ensure validity of the data at all levels: EDM, DM2E and vocabularies reused in EDM like Dublin Core. To avoid isolated solutions, DM2E initiated a task group at the Dublin Core Metadata Initiative<sup>8</sup> to address these questions together with Europeana and other stakeholders in the cultural heritage community. Starting point of the task group's work is the comparison of elements and validation rules in EDM, the DM2E model and the EDM application profile of the German digital library (DDB). The validation rules will be formalized and categorized. Application profile use cases from other projects and institutions will also be considered leading to a formulation of best-practices for the creation of RDF application profiles in general. Due to the central role that the DM2E model plays in this process, its awareness will be raised.

---

<sup>6</sup> <http://onto.dm2e.eu/>

<sup>7</sup> Application profiles like the DM2E model can specialise or refine vocabularies for a specific application, e.g. by mixing and matching certain ontology elements or by providing narrower element definitions.

<sup>8</sup> [http://wiki.dublincore.org/index.php/RDF\\_Application\\_Profiles](http://wiki.dublincore.org/index.php/RDF_Application_Profiles)

---

## 7 Documentation

### 7.1 Europeana Pro

Europeana has offered to host all deliverables and documentation on the Europeana Pro website, which ensures long-term and easy access to all Europeana partners. All of the documentation will also be available on the DM2E server at Humboldt-Universität.

### 7.2 Deliverables

All deliverables created by the project during the lifetime of the project are available on the Europeana Pro website: <http://pro.europeana.eu/web/digitised-manuscripts-to-europeana>. They are already available on the DM2E website. In order to continually make them accessible for the general public they will be transferred to the DM2E server hosted at Humboldt-Universität: <http://dm2e.eu/>.

### 7.3 DM2E Wiki

The DM2E Wiki, <http://wiki.dm2e.eu>, contains extended documentation and information on the developed infrastructure and software components as well as recommendations on mappings. The Wiki will be available on the DM2E server hosted at Humboldt-Universität.

The Wiki contains in detail:

#### **Information and documentation for Pundit and the Ask Learning Environment:**

- A user tutorials for in English, German, Italian and Norwegian
- Videos and screencasts
- A Pundit developer documentation for demonstrative different demonstrative applications (Edgemap, Timeline, Consuming data in Feed the Pundit, Solr based faceted search)
- A Pundit developer technical documentation on the Pundit server API, the server installation, the server source-code, the client user guide, the client Javascript API, client source code.

#### **An OmNom documentation consisting of**

- an English and German user manual, containing
  - an overview of the OmNom components,
  - the workflows in OmNom
  - information on how to use general functions, how to upload and manage files, how to create and edit workflows, how to configure workflows and how to configure and run jobs
- a screencast
- a walkthrough
- a technical documentation consisting of the
  - source code

- code documentation
- deployment of OmNom

### **The DM2E content provider mappings:**

This part is a resource for those who want to understand the specific mappings that DM2E content providers have used to convert their metadata to the DM2E data model. Information on the source formats used by each content provider is also included.

- Overview of the DM2E Model with the
  - DM2E model specification document
  - DM2E model OWL-file
  - Web representation of the DM2E model resources
- The mappings to the DM2E Data Model from the following source formats
  - MAB, MARC, TEI P5, METS/MODS, EAD, MPIWG internal, DC(X), ESE
- Tables of mappings from the following content providers
  - Providers listed in the description of work
    - Austrian National Library (ONB)
    - National Library of Israel (NLI)
    - Centre for Jewish History (CJH)
    - University of Bergen (UIB)
    - Humboldt-Universität zu Berlin (UBER)
    - Staatsbibliothek zu Berlin (SBB)
    - Berlin-Brandenburg Academy of Sciences (BBAW)
    - Max-Planck-Institut für Wissenschaftsgeschichte (MPIWG)
  - Associate providers joined during the project  
(documentation for providers who joined during the project is on a voluntary basis)
    - Universitätsbibliothek Frankfurt/Main (UBFFM)
    - Georg Eckert Institute for Textbook Research (GEI)
    - Eurocorr (ECORR)
    - The American Jewish Joint Distribution Committee (JDC)
    - Brandeis University (BRANDEIS)
    - Bulgarian Academy of Sciences (BAS)
    - Petőfi Literary Museum (PIM)
    - Koninklijke Bibliotheek (KB)
- Annotatable Content Specification with detailed technical information on the requirements for content to be used in the DM2E scholarly research platform (with tools like Pundit and Ask).
- Recommendations for the RDF representation of DM2E metadata with specific guidelines for content providers how to encode certain aspects of the data such as time information, URI design or the representation of subject terms and hierarchies.

### **Deployment documentation**

- of the other tools that are used in DM2E:

- deployment of Silk for data contextualisation
- deployment of Mint for data mapping

### SSO Documentation

- A technical documentation on the security service based on Java Open Single Sign-On framework (SSO) including
  - a SSO Introduction
  - Information about the gateway installation and agent configuration.

### Related open-source software

A list with a selection of open source tools of interest to DM2E, containing a description and information about developers, project website, code repository and lots more.

### Introduction to Open Cultural Data

This sections contains information on the basic concepts of open cultural data and how cultural institutions and individuals can take the first steps to opening up. It is based on material collected by the OpenGlam Working Group, which is part of the DM2E project.

## 7.4 Mapping Guidelines

The DM2E model is documented in different ways to support data providers in the project that create mappings from their formats to DM2E. The different documentation forms, pdf, owl and directly dereferenceable resources on the Web, are accessible for everyone and can be accessed via the DM2E wiki (see section 7.3).

The DM2E model specification in pdf format explains the model by providing definitions of reused resources and resources in the DM2E namespace, by giving the domain and range of properties and by showing original definitions of resources (if reused) as well as DM2E-specific scope notes. The provider can directly see whether elements were reused or not or if they part of the EDM or specialising it through different colour highlighting that was used in the document. Additional RDF examples at the end of every class description provide additional mapping support. The structure of the DM2E model specification is based on the Europeana Mapping Guidelines with which data providers may already be familiar with.

In addition to the model specification, mapping recommendations were created. They contain specific design-related recommendations which lead to homogenous mapping results.

## 7.5 DM2E Website

The DM2E website will be transferred from its current host to the DM2E server. Humboldt-Universität commits to keep this website running at least as long as the server. The transfer of the website and the domain name will be managed via OKFN.

---

## 7.6 Project Management Platform

The project management platform “Redmine” will be available on the DM2E server at Humboldt-Universität: <https://doc.dm2e.hu-berlin.de/redmine/>. All participants will have access to the internal project communication, documentation and files shared during the project term.

## 7.7 Github

The ontology and all source code developed in DM2E is available at <https://github.com/DM2E>.

While it would be feasible to clone the respective code source repositories on Europeana's Github (<https://github.com/europeana>), such localised space would most certainly not be active – certainly less than in their own community space. This is what Europeana has observed for a number of old contributions on its previous code source repository (<http://old.europeanalabs.eu/browser/contrib>).

### 7.7.1 Ontologies

The [dm2e-ontologies](#) repository on Github contains the ontologies created in DM2E and another ontology needed in DM2E, the EDM. DM2E ontologies are not only the DM2E model but also the OmNom vocabularies OmNom logging, OmNom types and OmNom ws that are used in the interoperability infrastructure. The DM2E validation tool, that validates DM2E data before it is ingested into the Triple store, as well as an automatic DM2E-to-EDM backwards mapping program are also accessible via the DM2E ontologies repository.

### 7.7.2 Scholarly research platform “Pundit”

Two of the core components, Pundit and Ask are already available on GitHub (<https://github.com/net7/pundit>, <https://github.com/simonefonda/ask-pundit>). Public Git repositories for Feed and Korbo will be put online before the end of the project. The code is actually under revision before being published.

### 7.7.3 OmNom

OmNom consists of two projects: the backend containing all the web services, triple store connections and helper libraries to cater for the various input formats in dm2e-ws while all the code and assets related to the web interface reside in dm2e-gui (with the JavaScript dependencies outsourced to a dedicated repository, `httpdm2e-gui-bower-deps`). Source Code documentation of OmNom is very complete, allowing developers to peruse the API documentation from within their IDE of choice or on the web (<https://doc.dm2e.hu-berlin.de/doc/dev-doc/apidocs/>). All the software developed except the JavaScript-based GUI and Bash command line scripts is written in Java and uses the industry standard Maven2 tool for building, deployment and dependency management.

---

#### 7.7.4 Validation Tools

Validation of RDF data against the DM2E model is provided as a set of Java classes that integrate the DM2E data model in dm2e-ontologies. For every new revision of the data model, a new version of the OWL ontology has been created with a corresponding Java class that models the updated constraints as programming code. The validation can be used in other Java projects using its API or as a stand-alone tool on the command line.

The EDM validation can be found in edm-validation git repository. It was created in collaboration with NTUA by extracting the relevant code from MINT. It is a two-step validation base on the RDF/XML serialisation of the data, consisting of an XML Schema step, validating basic constraints such as existence of certain triples, and a Schematron step, validating more advanced concepts such as conditional requirement of specific triples. Same as the DM2E validation, it can be used in Java code as well as on the command-line.

#### 7.8 Direct Ingestion Tools

For large-scale ingestions that make the use of OmNom impractical due to the large amount of overhead involved in its meticulous provenance tracking, WP2 implemented a command-line expert tool, residing in the [dm2e-direct git repository](#). This tool combines an optional XSLT transformation to DM2E with DM2E validation and ingestion into the triple store with basic provenance information. Interfacing parts of dm2e-ws and dm2e-ontologies in a straightforward command-line application, it is not only a very useful tool but can serve as a great entry point for future developers into DM2E's code base.

#### 7.9 DM2E Analyses

Using the code in [dm2e-analysis](#) git repository, DM2E has been quantitatively analysing and visualising the provided data. Fine-tuned SPARQL queries extract the relevant information from the triple store, the resulting tabular data is processed by a Python script and the results are written as a set of HTML/JavaScript files to be visualised using Google Charts. While not as exhaustive documented as other parts of the code base, it can help other Linked Data developers to evaluate the SPARQL queries and understand the structure of the triple store.

#### 7.10 Pubby

Pubby is the mediator that exposes the RDF data stored in the triple stores for the DM2E data and the OmNom application data to the world using an HTTP API and a user interface on top of that. While originally a widely used and appreciated tool in itself, the DM2E fork of Pubby offers additional features and its configuration file is probably the best place to understand the DM2E data since it not only explains the structure of the data in the triple store by means of SPARQL queries but also shows its limitations and parts where the queries are optimised to balance performance and recall.



---

## 7.11 Vimeo

The videos created by the DM2E project will remain at the video hosting platform Vimeo. The DM2E Vimeo account can be found at <http://vimeo.com/user19973958>. All videos support community annotation.

---

## 8 Community

One of the primary objectives of DM2E's dedicated dissemination work package has been to build a community of open metadata evangelists. A vibrant, sustainable community of individuals who are able to evangelise about the benefits of open data for cultural institutions is critical for the growth of the Linked Open Web which the DM2E tools are built upon.

### 8.1 Open Glam

The community of open metadata evangelists which has been established by DM2E was specifically branded as the "OpenGLAM" community to allow for it to continue to exist beyond the lifetime of the project and to give it sufficiently broad appeal.

The OpenGLAM community is established in such a way that it can be led by volunteers without the need for central funding. The DM2E project has invested significant resources in establishing the community structures that will allow the network of open data evangelists to expand beyond the lifetime of the DM2E project. Their key community structures are:

- A volunteer led Working Group led by a volunteer coordinator with a process for membership;
- A high-profile Advisory Board of thought leaders in the field of open cultural data;
- Documentation and evangelism material including slide decks and handbooks on open cultural data;
- Stocks of publicity material including stickers, flyers and posters for those wanting to promote open metadata at events and conferences;
- The establishment of OpenGLAM local groups as part of Open Knowledge Foundation national chapters. The [Finnish OpenGLAM local group](#) has already secured funding to promote open cultural data to Finnish cultural heritage institutions with the support of the international OpenGLAM Working Group. Additional local groups have been set up in [Austria](#) and [Switzerland](#), with a group in [Germany](#) starting in 2014.

The Open Knowledge Foundation is committed to supporting this community and area of activity as part of its global network of thematic Working Groups. It will continue to host OpenGLAM.org and all other web based community infrastructure such as the OpenGLAM discussion and mailing lists.

In addition, there is a dedicated Working Group coordinator at the Open Knowledge Foundation who will continue to support the OpenGLAM Working Group along with the other thematic Working Groups on open data at the Open Knowledge Foundation.

### 8.2 Open Humanities Awards

One of the key outcomes of work package 4 has been to create the "Open Humanities Awards", a contest that supports open source innovation based on open humanities data. A dedicated website was set up (<http://openhumanitiesawards.org>) and the awards were published widely across all major Digital Humanities lists and channels. With over 50 applicants in the first round, and over 20 applications in the second round, the awards



---

have been a success so far, which is why the Open Knowledge Foundation aims to continue hosting this web platform and hold new rounds of the Open Humanities Awards in the future.

---

## 9 Appendix A: DM2E Proposal for Extending the EDM Implementation of Europeana

**Authors:** Doron Goldfarb (ONB)

**Contributors:** Kai Eckert (UMA)  
Antoine Isaac (Europeana)  
Christian Morbidoni (Net7)

### Abstract

This part of D5.10 proposes extensions to Europeana's EDM implementation based on requirements identified throughout the course of the DM2E project. The underlying rationale is to transfer the DM2E specific metadata representation of annotatable text/image content into a form that can be hosted by Europeana without relying on specialised DM2E data infrastructure. This proposal is divided in three sections. Section 9.1 discusses how Web Resources representing parts – in the case of DM2E pages or paragraphs of a manuscript – of the content of a Cultural Heritage Object (CHO) can be attached to the metadata of that CHO and thus made accessible to DM2E's annotation software. Section 9.2 discusses how to extend the object display in the Europeana portal by providing a link to DM2E's annotation software so that Europeana visitors are enabled to annotate the content representations discussed in section 9.1. Section 9.3 concludes with a brief discussion of necessary changes for DM2E's annotation software. Section 9.4 concludes with a short statement on the current state of discussion with Europeana regarding the proposed changes.

## 9.1 Representing DM2E annotatable content in EDM

This section presents the current metadata representation of DM2E annotatable content within the project's data infrastructure and discusses issues with providing the same functionality in Europeana's EDM implementation. Based on the discussion, a number of proposals to the EDM are made in order to enable content provision to Europeana that includes features such as document annotation.

The underlying rationale is to create means to provide metadata to Europeana that do not necessarily have to be displayed in the Europeana portal but can nevertheless be retrieved through the Europeana API. By "decoupling" certain aspects of Europeana EDM metadata from their immediate display in the Europeana portal, third party services such as the DM2E annotation software could access relevant metadata directly through the Europeana API without interfering with the portal. In the case discussed in this document, this mainly applies to Web Resources attached to Aggregations.

### 9.1.1 Introduction

During the DM2E project, a number of content providers deliver metadata about digitised manuscripts and other bibliographic items to Europeana. One significant feature of DM2E's contribution is making the digitised content behind some of the provided metadata records accessible to annotation software such as the tool Pundit which is being developed as part of the DM2E project.

The metadata structure for such annotatable content depends on the representation of the respective digitised items on the providers Website. In many cases, the content is accessible in the form of digitised facsimiles representing the pages of the provided cultural heritage object. In DM2E, the approach to representing the structure of such documents is to model it as a hierarchy of ProvidedCHOs, where the level of finest granularity can be arbitrarily chosen. Usually, this level is composed of ProvidedCHOs representing pages. Figure 1 shows such an example for a multivolume book title. Given an entry point which usually is the ProvidedCHO representing the full title as returned by a search result, the annotation software accesses this hierarchy through DM2E's data infrastructure in order to identify the annotatable elements.

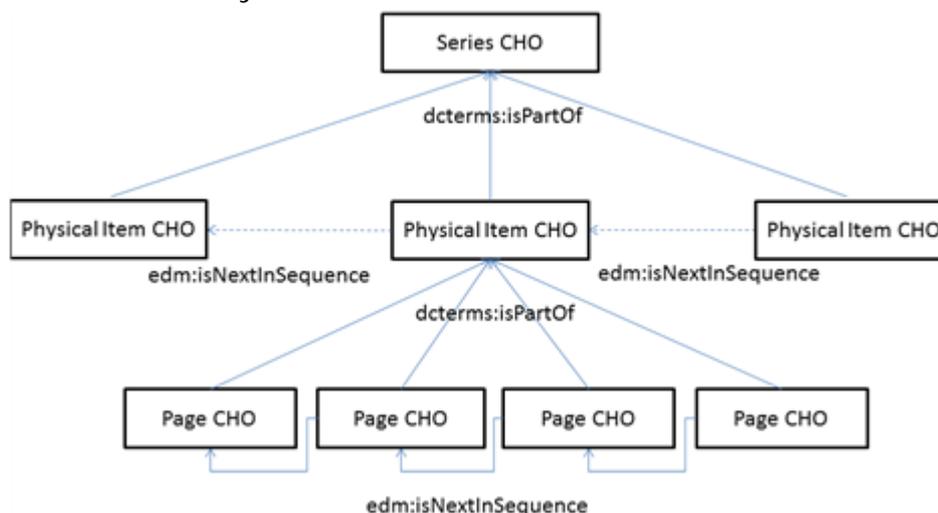


Figure 1: Representation of DM2E annotatable content

---

This fine-grained access is currently provided through DM2E's triple-store and linked data API. As the metadata records with DM2E content will also be ingested to Europeana throughout the project, it seems natural to try to represent them in Europeana in a way that makes it possible for the Pundit annotation software to annotate the underlying content directly through Europeana's own metadata API.

### 9.1.2 Providing DM2E Object Metadata to Europeana

The DM2E project will publish its data to Europeana during DM2E's lifespan, using Europeana's existing ingestion process. As this is already covered in the subcontract between DM2E and the Europeana Foundation, updates will be delivered during the duration of the project by DM2E as aggregator for all data providers in the framework of the DM2E project.

The DM2E project aggregates data from providers using an extension of the EDM, the DM2E data model. Europeana is not able to aggregate this extension of the EDM yet and will therefore ingest the 'basic' EDM. DM2E will convert its RDF data to fit the constraints set in EDM/XML and will transfer it to Europeana via OAI-PMH.

In principle, the DM2E hierarchy of ProvidedCHOs fully conforms to the EDM model. The reason why it is not possible to ingest the hierarchy "as-is" into Europeana lies in Europeana's current policy regarding the granularity of provided objects for its own environment. It states correspondingly that as long as there are no specific metadata available describing significant properties of single pages belonging to a superordinate object, they are not considered as being informative enough to be represented as Europeana objects themselves.

Thus, the DM2E page-level CHOs are discarded during the mapping from the DM2E model to EDM. This has the effect that it will not be possible to annotate DM2E objects (single pages) based on their Europeana metadata representation alone. Therefore this proposal seeks to identify possibilities to represent annotatable DM2E content in EDM in a form that

1. is compatible with EDM
2. fulfils Europeana's requirements regarding the granularity of provided objects
3. is compatible with DM2E's annotation software Pundit

### 9.1.3 Alternative Representation of Low-level, Annotatable Content (= Pages)

In order to fulfil requirements 1 and 2 from the list above, the annotatable content has to be represented in a way that is compatible with EDM and does not produce "additional" metadata records that do not fulfil Europeana's requirements. This effectively means that pages are not allowed to be represented as separate ProvidedCHO and that thus an alternative way of their representation has to be sought.

As an alternative, the pages of a manuscript could be connected to the ore:Aggregation of the respective CHO as edm:WebResources. The Web resources can be put into sequence by their edm:isNextInSequence properties and connected to the ore:Aggregation by edm:hasView. This solution basically fulfils requirements 1 and 2, allowing the representation of pages without interfering with Europeana's guidelines for provided objects. In principle it also meets requirement 3 in that it would be possible to adapt

Pundit content access routines to deal with this alternative presentation. Figure 2 shows a sketch of the alternative metadata structure.

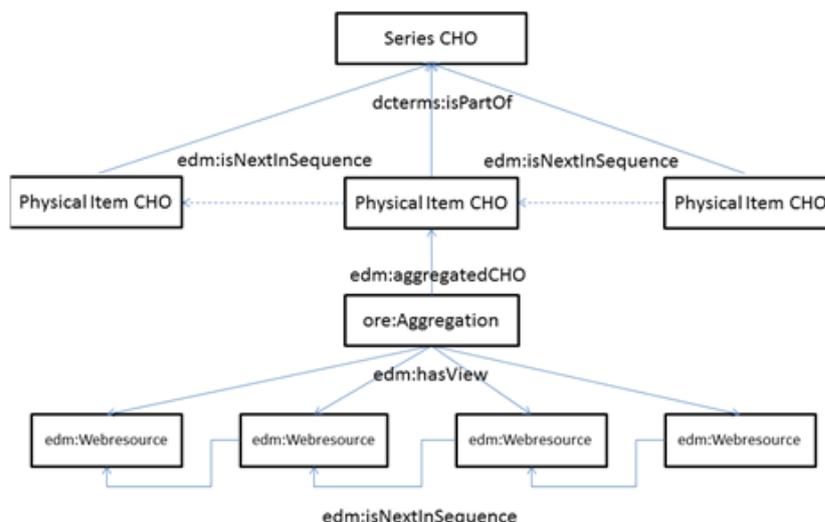


Figure 2: Representing pages as edm:WebResource connected through edm:hasView

### 9.1.3.1 Disadvantages of Current EDM Implementation

There are, however, two main disadvantages that have to be taken into account:

#### **Web resources attached to an ore:Aggregation through edm:hasView show up in the respective image carousel/lightbox display in the Europeana portal**

Sometimes providers might like to attach “content views” to the data that should be available through the Europeana API but not be visible in the Europeana portal.

#### **edm:hasView is a generic connection between a Web resource and an ore:Aggregation.**

Consumers of this property can’t distinguish between different Web resources attached through edm:hasView besides checking their respective “dc:format” property, when available. A parallel presence of e.g. different digital image files of which only a subset is designated to be annotatable would cause problems for an annotation software to decide which Web resources should be accessed. This could, of course, also be addressed by the data provider through specific guidelines what to include in edm:hasView links.

### 9.1.3.2 Proposal for two EDM Extensions

In order to address these two potential shortcomings, this proposal raises two suggestions for the extension of the EDM. Both are targeted to providing additional means to attach Web resources representing parts of the content of a provided object to its ore:Aggregation such that the attached Web resources do not necessarily show up in the Europeana portal. Especially in regard of the planned transition “from portal to platform” this would enable data representations of content links independent from the Europeana portal display, allowing third party data consumers to offer additional services not interfering with Europeana’s interpretation of the EDM.

The two proposed extensions are listed below:

**a) Provide an alternative property for ore:Aggregation that is similar to edm:hasView but is ignored by the portal**

The first extension would leave edm:hasView as property for including Web resources in the Europeana portal display without introducing additional interpretations. Such “special” uses for Web resources attached to aggregations could be catered through an additional property such as “*edm:hasContentWebResource*”. It could also be a sub-property of hasView which the Europeana display may filter out.

**b) Add dc:type property to edm:WebResource**

The second proposed extension is to add the property “*dc:type*” to the class **edm:WebResource**. This would enable data providers to explicitly encode the intended use of the Web resource in question for machine access. This could also be accompanied by a controlled vocabulary, providing terms such as “annotatable content”, etc.

### 9.1.3.3 Discussion

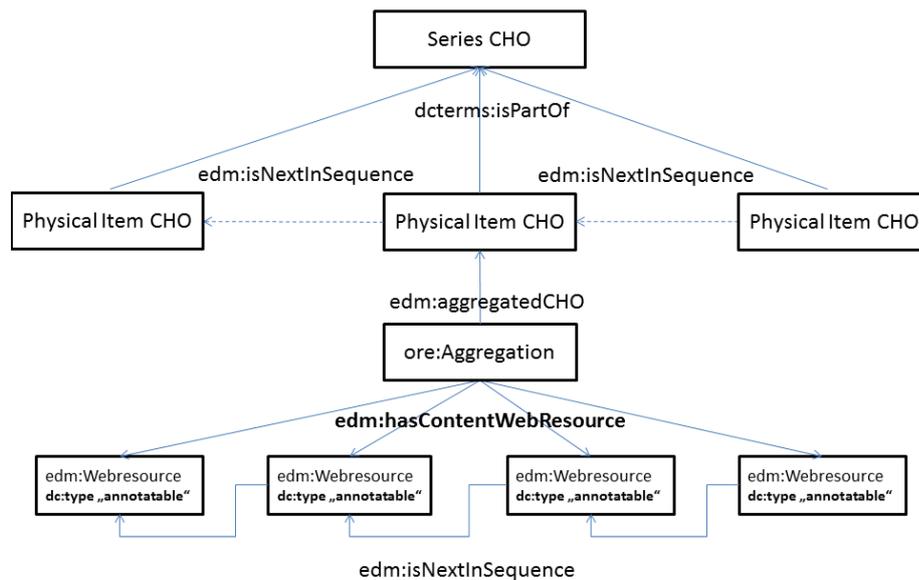


Figure 3: Representing pages as edm:WebResource including the proposed EDM extensions

Figure 3 shows the metadata structure including the proposed EDM extensions. This way it will be possible for data providers to connect Web resources to Aggregations using **edm:hasContentWebResource** without interfering with the Europeana portal display. Through the additional **dc:type** property, each WebResource could moreover be “typed” with respect to their designated usage.

### Differences and Similarities between the two Extensions

In principle each of the two proposed extensions alone would be sufficient to fulfil the basic requirement to be able to attach Web Resources to Aggregations without “interfering” with the Europeana portal. The difference lies mainly in the flexibility of the proposed extensions and their implications on necessary modifications in the portal software.

Extension 9.1.3.2 a) has the advantage that providers could directly use the newly created alternative property to `edm:hasView` for attaching Web Resources to Aggregations without causing “effects” in the Europeana portal. No further changes are needed for the portal.

Extension 9.1.3.2 b) on the other hand allows more flexibility through typed Web Resources, potentially extending the functionality that is provided by adding one alternative property. The realization of this more flexible solution would, however, cost more resources, as appropriate filters would have to be implemented in the Europeana portal software and a controlled vocabulary or type system should be provided for Web Resource types.

## 9.2 Providing a Link to the Annotation Software

### 9.2.1 Introduction

While section 9.1 of this document discusses the provision of “low-level” digital representations of an object such as e.g. Web Resources linking to single page scans of books and issues with their attachment to the object’s aggregation, this section focuses on “high-level” representations of digital objects instead. “High-level” object representations refer to Web Resources linking to “whole” object views such as provided through the EDM properties *isShownAt* and *isShownBy*. In the Europeana portal, these properties are dedicated to linking to whole object views on the original providers’ Websites. No means to link to other/third party representations of digitised object views are currently provided in the portal.

### 9.2.2 Providing means to link from the Europeana portal to external representations

In the context of DM2E, an additional link to an alternative Web Resource would be necessary to link from the Europeana portal representation of DM2E contributed objects to DM2E’s annotation software. As explained in section 9.1.1 and shown in figures 1-3, the annotation software needs an entry point for traversing all the annotatable Web Resources belonging to a contributed object. This entry point usually is the metadata record that represents the “whole” ProvidedCHO such as the whole manuscript and it can be provided through the Europeana API as RDF.

DM2E’s annotation software is a Web application callable through a dedicated URL that looks as follows:

```
http://feed.thepund.it/?dm2e=[DEREFERENCEABLE_URI_OF_OBJ_METADATA_REPR]&conf=d  
m2e.js
```

Given that it will be possible to represent annotatable content metadata as outlined in section 9.1.3, it would therefore be desirable to embed links of the form

```
http://feed.thepund.it/?dm2e=[EUROPEANA_API_CALL_TO_OBJECT_METADATA_RDF]&conf=  
dm2e.js
```

in the Europeana portal display of annotatable DM2E objects. This section discusses the current options to provide such links and discusses their disadvantages. Based on this discussion two additional EDM extensions are proposed.

### 9.2.2.1 Disadvantages of current options to provide additional external links

Currently there are proposed workarounds suggesting to link to the respective Web resources through the edm:ProvidedCHO properties **dc:source** or **edm:isRepresentationOf**. The problem with these solutions is twofold:

#### Violation of EDM Specification

If one uses dc:source or edm:isRepresentationOf then a connection is made between a ProvidedCHO and a Web resource, which according to the EDM definition should actually be done through ore:Aggregation.

#### No Context Information for External Links

Since dc:source and edm:isRepresentationOf are rather generic properties, an attached Web resource would show up as URL in the portal without any additional context information on the purpose or function of that URL.

#### Example



Figure 4: Adding additional external links to the item view in Europeana. dc:source workaround

While issue 1 (violation of EDM specification) is rather conceptual, issue 2 (no context information for external links) makes it inconvenient to provide external URLs this way. Figure 4 shows an example view in the Europeana portal. The link to the DM2E representation of the CHO is provided through dc:source. No information can be added to the “raw” display of the link. This is disappointing, as for example clearly described links to “third-party” services such as annotation software could be of added value to Europeana visitors. In the case of DM2E, this would apply to links to DM2E’s annotation software Pundit that could be embedded into the Europeana portal view.

### 9.2.2.2 Proposed Solution

This section makes an additional proposal to enable the inclusion of such “third-party” links into Europeana in a way that is more pleasing than the currently proposed workarounds suggest. The proposal is twofold, addressing the display of Web resource URLs in general (issue 2 above) and also mentioning the problem of providing Web resource links in a provided CHO instead of an aggregation (issue 1):

## a) Labels for Web Resources

For a Web resource connected to a property visible in the Europeana portal display, we propose to enable providers to add a Web resource-label that will be displayed in the Europeana portal in place of the URL. The hyperlink in the portal would then look (somewhat) as follows:

```
<a href="http://feed.thepund.it/?dm2e=http://data.dm2e.eu/data/item/sb  
b/manumed/PPN620004592&conf=dm2e.js">Annotate &quot;De origine  
comoediarum - Petrus Guntherus&quot; with Pundit</a>
```

We think about two different options for providing this label:

1. **Provide the label through the Web resource's dc:description property.** Wherever a property visible in the Europeana display is attached to a Web resource and dc:description of the Web resource is not empty, display the literal in dc:description the way described above
2. **Extend edm:WebResource by the additional property skos:prefLabel** (or dc:title) and use it instead of dc:description as described in a)

## b) Additional Property in Aggregation Class

We suggest to provide an additional property such as edm:externalLink in ore:Aggregation alongside *isShownBy* and *isShownAt*. This extension would allow to attach additional Web resources for a given CHO in line with the EDM specification by attaching the Web resource via the aggregation. Moreover, compared to dc:source or edm:isRepresentationOf this provides a dedicated property for additional external Web resources relevant to the respective CHO.

### 9.2.2.3 Discussion

The implementation of the extensions "Labels for Web Resources" and "Additional property in Aggregation class" for attached Web resources will allow to provide meaningful links to DM2E's annotation facilities directly from Europeana. Figure 5 shows a mock up for the proposed solution as we would imagine it showing up in the Europeana portal.



Public Domain Mark - "No Known Copy right"

View item at Austrian National Library

Share

Cite on Wikipedia

Translate details

Select language

Powered by Microsoft Translator

**External link: Annotate "Collectio theologica" with Pundit**

## Collectio theologica

**Description:** Cod. Theol. gr. 134 HAN ; Hofbibliothekseinband mit Supralibros des Gerard van Swieten, datiert 1755; Restaurierung im Jahre 1918; Papier; Zierlinien und Initialen in brauner Federzeichnung, Bandinitialen, Überschrift und Initiale Pi in karmesinroter Tinte, Zopfband in brauner Federzeichnung; Bl. 1r-86v: Basilius Caesariensis, Hom. in Hexaemeron (inc. mut.) Hom. 1-9, Hom. 1 de creatione hominis+ (Bl. 87r-94r), Constitutiones monasticae Cap. 2.17.3-5-7-13-16.18. 22-23.27-29.11.19.6.8-10.20-21.24-26.30-34 (Bl. 151r-182v); Bl. 10v: Gregorius Nyssenus, In Hexaemeron (Excerpta), De opificio hominis cum Scholiis Cap. 1-30(Bl. 95r-150r); Bl. 150v: Theodorus Studita, Scholia in Basili Ascetica; Bl. 182v-183r: Johannes Climacus, Liber 1 (Excerpta); Bl. 183r-184r: Anastasius Sinaita, Viae dux Cap. 1 (Excerpta); Bl. 185r.209v -211rv.210rv.212r.220rv.214r-219v.213rv.221r-237r: Eustathius Thessalonicensis, De emendanda vita monastica; Bl. 212v.220rv.214r-215r: Excerpta ex operibus de medicina; Bl. 238r-262v: Theodorus Smyrnaeus, De natura et principiis naturalibus cum Scholiis Cap. 1-3; Bl. 263r-270v: Antonius Studita, Hom. de Acathesto

**Creator:** <http://d-nb.info/gnd/10093630X>; <http://d-nb.info/gnd/102407681>; <http://d-nb.info/gnd/118541919>; <http://d-nb.info/gnd/118558048>; <http://d-nb.info/gnd/118637797>; <http://d-nb.info/gnd/118682733>; <http://d-nb.info/gnd/118756761>; [http://data.dm2e.eu/data/agent/onb/codices/Antonius\\_Studita](http://data.dm2e.eu/data/agent/onb/codices/Antonius_Studita); [http://data.dm2e.eu/data/agent/onb/codices/Bacchus\\_Levita](http://data.dm2e.eu/data/agent/onb/codices/Bacchus_Levita)

**Publication date:** [http://data.dm2e.eu/data/timespan/onb/codices/1200-01-01T000000UG\\_1249-12-31T235959UG](http://data.dm2e.eu/data/timespan/onb/codices/1200-01-01T000000UG_1249-12-31T235959UG); 13. Jahrhundert, 1. Hälfte

**Type:** <http://onto.dm2e.eu/schemas/dm2e/Manuscript>

**Format:** 260/265 x 160/175 mm; Handschrift; I+271 Bl.

**Identifier:** +Z10395630X

**Relation:** Manuscript

**Language:** grc

**Search also for:**

**Title**  
Collectio theologica (1)

**Who**  
<http://d-nb.info/gnd/10093630X> (1)  
<http://d-nb.info/gnd/102407681> (1)  
<http://d-nb.info/gnd/118541919> (1)  
<http://d-nb.info/gnd/118558048> (1)  
<http://d-nb.info/gnd/118637797> (1)  
<http://d-nb.info/gnd/118682733> (1)  
<http://d-nb.info/gnd/118756761> (1)  
[http://data.dm2e.eu/data/agent/onb/codices/Antonius\\_Studita](http://data.dm2e.eu/data/agent/onb/codices/Antonius_Studita) (1)  
[http://data.dm2e.eu/data/agent/onb/codices/Bacchus\\_Levita](http://data.dm2e.eu/data/agent/onb/codices/Bacchus_Levita) (1)

**What**  
<http://onto.dm2e.eu/schemas/dm2e/Manuscript> (1) (6)

**Provider**  
Austrian National Library (512662)  
DM2E (34)

Figure 5: Mock up for proposed solution to include external links in ore:Aggregation including a label for the attached Web resource

### 9.3 Necessary Changes for the DM2E Annotation Service

The DM2E annotation platform currently uses the DM2E data representation. In order to work with the proposed EDM extensions, the data processing module of the DM2E software will have to be updated accordingly. The main differences will be:

The data will be retrieved through the Europeana API/Linked Data interface. Links to the DM2E annotation service embedded in the Europeana portal as described above will have to be changed:

```
<a href="http://feed.thepund.it/?dm2e=http://data.dm2e.eu/data/item/sbb/manumed/PPN620004592&conf=dm2e.js">Annotate "De origine comoediarum - Petrus Guntherus" with Pundit</a>
```

will become

```
<a href="http://feed.thepund.it/?dm2e=LINK_TO EUROPEANA DATA INTERFACE&conf=dm2e.js">Annotate "De origine comoediarum - Petrus Guntherus" with Pundit</a>
```

One potential issue here would be the API key that has to be included when using the Europeana API. By embedding a link including the API call, the Europeana API key would be disclosed. Therefore it would be necessary to get access to the Europeana API without

---

an API key. Ideally, access to the Europeana RDF data will be possible in a pure Linked Data fashion, without having to use an API key at all.

The data will be represented in the Europeana “flavor” of EDM. The data processing module of the DM2E annotation platform will have to be able to parse the Europeana RDF output, including the proposed extensions. Currently, the RDF output of the Europeana API is limited to RDF+XML or JSON-LD.

## 9.4 Current state of discussion with Europeana

This section gives an overview on the current state of the ongoing discussion between DM2E and Europeana regarding the extensions proposed in sections 9.1.3 and 9.2.2.

### 9.4.1 Outcome of Discussion between Europeana Product Development and R&D regarding proposed changes section 9.1.3

In general, Europeana currently rejects the proposed specification of an additional property to link page representations to provided cultural heritage objects and refers to the existing property `edm:hasView`. They argue that everything that is deemed annotatable by providers should also be displayed in the Europeana portal. As long as the portal is able to deal with displaying huge amounts of page-WebResources, DM2E could in principle use the existing `edm:hasView` property and waive the proposal from section 9.1.3, although at the cost of losing flexibility for providers: In general, it would be appreciated to be able to provide specific data about an object through the Europeana API without having to display them in the Europeana portal.

### 9.4.2 Outcome of Discussion between Europeana Product Development and R&D regarding proposed changes section 9.2.2

According to Europeana, DM2E’s proposal to add a flexible mechanism for providers to include links to additional services is not feasible due to potential misuse and resulting quality issues in the Europeana portal. Europeana made some alternative suggestions that will have to be discussed further. In contrast to the proposal in section 9.1.3, the requirements raised in section 9.2.2 are crucial for providing Europeana visitors with the opportunity to call DM2E’s annotation platform directly from the Europeana portal for suitable DM2E content.

## 10 Appendix B: Follow-Up Aggregators

Europeana is assisting DM2E in finding aggregators for the DM2E providers. The following table represents the status of July 30th, 2014 according to Europeana and is not completed.

Provider	Country	Possible Future Aggregator	No. of Pages	Type	Notes
<b>SBB</b> Staatsbibliothek zu Berlin	Germany	TEL / DDB	206000	Manuscripts and personal papers	Are partner in TEL
<b>MPIWG</b> Max-Planck-Institut für Wissenschaftsgeschichte	Germany	DDB	640000	Manuscripts, books and autographs	
<b>BBAW</b> Berlin-Brandenburgische Akademie der Wissenschaften	Germany	DDB	418000	Books and historical printings	
<b>UBER</b> Humboldt-Universität zu Berlin	Germany	DDB	200000	Scientific Journal	Are also a partner in BHL
<b>UBFFM</b> Universitätsbibliothek Frankfurt am Main	Germany	DDB	238000	Manuscripts	Are also a partner in Judaica and Europeana Local Germany
<b>ONB</b> Austrian National Library	Austria	TEL	15640000	Books, periodicals and manuscripts	Are partner in TEL
<b>UIB</b> Universitetet i Bergen	Norway	Arts Council Norway	5000		Are already a partner in Arts Council Norway
<b>BAS</b> Bulgarian Academy of Sciences	Bulgaria	<a href="http://bulgariana.eu">http://bulgariana.eu</a>	570	Manuscripts	Are also a partner in Athena and Linked Heritage
<b>ECORR</b> "The European Correspondence to Jacob Burckhardt"	Italy	APEX	2800	Manuscripts	



---

Provider	Country	Possible Future Aggregator	No. of Pages	Type	Notes
<b>NLI</b> National Library of Israel	Israel	Judaica?	2650000	Manuscripts, books and archival items	
<b>CJH</b> Center for Jewish History	United States	Judaica?	400000	Books and photographs	
<b>JDC</b> American Jewish Joint Distribution Committee	United States	Judaica?	20000	Archival items	
<b>BRANDEIS</b> Goldfarb Library Brandeis University	United States		650	Handwritings	